

#####

#. PANEL DATA ANALYSIS

#

#####

#1. Introduction of Panel Data Analysis

Panel data, also known as longitudinal data, combines elements of time series and cross-sectional data,

#providing a multifaceted perspective on changes over time among various entities.

#It identifies trends and accounts for individual-specific effects by employing fixed or random effects

#models to address unobserved heterogeneity.

#This type of data enhances the understanding of temporal relationships and supports informed decision-making.

#Techniques such as dynamic panel data models offer deeper insights into complex economic analyses. Further exploration reveals its intricate features.

#Panel data combines time series and cross-sectional data, offering insights into dynamic changes over time across multiple entities.

#It facilitates analysis of dynamic variables, revealing trends and patterns overlooked by single-dimension data.

#Fixed and random effects models address unobserved heterogeneity for accurate economic insights.

#Homogeneous models apply uniform parameters, while heterogeneous models account for individual-specific variations.

#Dynamic panel data models manage autocorrelation and unobserved heterogeneity through advanced techniques like Arellano-Bond estimators.

#3.Understanding the Concept of Panel Data

#Panel data, often referred to as longitudinal data, plays an essential role in econometrics by offering a detailed view of dynamic

#changes over time across multiple entities, such as individuals or countries.

#It combines the characteristics of time series and cross-sectional data, enabling an in-depth economic analysis. By observing multiple

#cross-sectional units, panel datasets reveal trends and patterns that single-dimension data might miss.

#Fixed effects and random effects models further enrich this analysis by addressing unobserved heterogeneity, ensuring more accurate insights.

#Whether balanced or unbalanced, these datasets empower researchers to better understand and serve societal needs through informed decision-making.

#Characteristics and Structure of Panel Datasets

#1. Building on the foundational understanding of panel data, it is crucial to comprehend the specific characteristics and structure these datasets embody.

#2. Panel data consists of multiple observations across time periods for the same cross-sectional units, enabling analysis of dynamic variables.

#This format allows identification of individual-specific effects that may influence outcomes.

#3. Datasets can be balanced, with equal observations for each unit, or unbalanced, with missing data. Presented in long format,

#panel data facilitates econometric analysis by addressing correlation within groups over time.

#This structure improves understanding of temporal relationships, providing insights unattainable with purely cross-sectional or time series data.

#Exploring Homogeneous and Heterogeneous Models

#How do economists decide which model to use when analyzing panel data? The choice hinges on the presence of individual-specific effects and desired analysis depth.

#Homogeneous models, like pooled OLS, apply uniform parameters, often neglecting variability in individual-specific effects.

#In contrast, heterogeneous models—fixed effects and random effects—embrace this variation.

#Fixed effects models capture individual-specific effects, assuming they correlate with observed characteristics, while random effects models,

#assuming no correlation, analyze both between and within-individual variation.

#Two-Way Individual Effects Models further improve analysis by integrating time-specific effects, offering a thorough econometric perspective on panel data.

#Analyzing Individual-Specific Effects in Panel Data

#Understanding individual-specific effects is essential for economists when analyzing panel data, as these effects represent unobservable

#characteristics that vary across individuals but remain constant over time.

#Econometric models effectively capture these effects:

#(a) The One-Way Fixed Effects Model controls for unobserved heterogeneity by correlating with observed variables.

#(b) The One-Way Random Effects Model assumes uncorrelated individual-specific effects, analyzing both within and between-group variations.

#(c) Two-Way Individual Effects Models incorporate time-specific effects, capturing unobserved factors across dimensions.

#(d) Dynamic Panel Data Models address autocorrelation by including lagged dependent variables, revealing deeper temporal relationships.

```
library(foreign)
```

```
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
```

```
#Data Structure
```

```
head(Panel) # viewing the snap short of the panel data (first six observation)
```

```
View(Panel) #viewing the full data set (lets discuss the  
variables/features/covariates/explanatory,predictors)
```

```
str(Panel) # Structure of each variable
```

#Exploratory Data Analysis

```
coplot(y ~ year|country, type="l", data=Panel)    # Lines (dependent by year per country)
```

```
coplot(y ~ year|country, type="b", data=Panel)    # Points and lines (dependent by year per country)
```

```
#library(foreign)
```

```
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
```

```
library(car)
```

```
#scatterplot(y~year|country, boxplots=FALSE, smooth=TRUE, reg.line=FALSE, data=Panel)
```

#FIXED-EFFECTS MODEL

```
#library(foreign)
```

```
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
```

```
library(gplots)
```

```
# Plot of means dependent against country
```

```
plotmeans(y ~ country, main="Heterogeineity across countries", data=Panel)
```

```
# plotmeans draw a 95% confidence interval around the means
```

```
#library(foreign)
```

```
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
```

```
#library(gplots)
```

```
# Plot of means dependent against year
```

```
plotmeans(y ~ year, main="Heterogeineity across years", data=Panel)
```

```

#library(foreign)
Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")

#Fitting the ordinary least squares model(linear model) (assuming normality of the
dependent variable)
ols <- lm(y ~ x1, data=Panel)
summary(ols)

#Comment on the X1 (Statistically insignificant)

#Regular OLS regression does not consider heterogeneity across groups or time
yhat <- ols$fitted

plot(Panel$x1, Panel$y, pch=19, xlab="x1", ylab="y")
abline(lm(Panel$y~Panel$x1),lwd=3, col="red")

#Fixed effects using Least squares dummy variable model
#library(foreign)
#Panel <- read.dta("http://dss.princeton.edu/training/Panel101.dta")
fixed.dum <- lm(y ~ x1 + factor(country) - 1, data=Panel)
summary(fixed.dum)

#Comment on the X1 (Statistically significant)

#Least squares dummy variable model
library(car)
scatterplot(yhat~Panel$x1|Panel$country, boxplots=FALSE, xlab="x1",
ylab="yhat",smooth=FALSE)
abline(lm(Panel$y~Panel$x1),lwd=3, col="red")

```

```

library(stargazer)
stargazer(ols, fixed.dum,
  type = "text", # Change to "html" or "latex" if needed
  title = "Comparison of Two Models",
  align = TRUE,
  digits = 2,
  out = "model_comparison.txt") # Save output to a file if desired

#Comment on OLS:The coefficient of x1 indicates how much Y changes when X
increases by one unit.

#Notice x1 is not significant in the OLS model

#Comment The coefficient of x1 indicates how much Y changes overtime,
controlling by

#differences in countries, when X increases by one unit. Notice x1 is significant in
the LSDV model

#Fixed effects: n entity-specific intercepts (using plm)
library(plm)
fixed<-plm(y ~ x1, data=Panel, index=c("country", "year"), model="within")

#Notes
#1. x-Outcome variable
#2. x-preictor varoable
#3. Panel-data
#4. Panel settings-country, year
#5. within-fixed effects option

summary(fixed)

```

#The coeff of x1 indicates how much Y changes overtime, on average per country, when X increases by one unit.

#P-value corresponding to x1: $\Pr(>|t|)$ = Two-tail p-values test the hypothesis that each coefficient is different from 0.

#To reject this, the p-value has to be lower than 0.05 (95%, you could choose also an alpha of 0.10),

#if this is the case then you can say that the variable has a significant influence on your dependent variable (y).

#If this number is < 0.05 then your model is ok. This is a test (F) to see whether all the coefficients in the model are different than zero.

Display the fixed effects (constants for each country)

fixef(fixed)

Testing for fixed effects, null: OLS better than fixed

#If the p-value is < 0.05 then the fixed effects model is a better choice

pFtest(fixed, ols)

#RANDOM-EFFECTS MODEL

random <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="random")

summary(random)

#To decide between fixed or random effects you can run a Hausman test where the

#null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects

phtest(fixed, random)

#If the p-value is significant (for example <0.05) then use fixed effects, if not use random effects.

```
#####  
#####  
#
```

#DIAGNOSTICS

#1. Testing for time-fixed effects

```
library(plm)
```

```
pool <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="pooling")
```

```
summary(pool)
```

#The null hypothesis in the LM test is that variances across entities is zero.

#This is, no significant difference across units (i.e. no panel effect)

```
plmtest(pool, type=c("bp"))
```

#Here we failed to reject the null and conclude that random effects is not appropriate.

#This is, no evidence of significant differences across countries, therefore you can run a simple OLS regression.

#The null is that no time-fixed effects needed

```
pFtest(fixed.time, fixed)
```

```
plmtest(fixed, c("time"), type=("bp"))
```

```
#####
```

#2. Testing for random effects: Breusch-Pagan Lagrange multiplier (LM)

```
pool <- plm(y ~ x1, data=Panel, index=c("country", "year"), model="pooling")
```



```
summary(pool)
```

```
#The LM test helps you decide between a random effects regression and a simple OLS regression.
```

```
#
```

```
#####
```

```
#3. Testing for cross-sectional dependence/contemporaneous correlation: using Breusch-Pagan LM test of independence and Pasaran CD test
```

```
#cross-sectional dependence is a problem in macro panels with long time series.
```

```
#The null hypothesis in the B-P/LM and Pasaran CD tests of independence is that residuals across entities are not correlated.
```

```
#B-P/LM and Pasaran CD (cross-sectional dependence) tests are used to test whether the residuals are correlated across entities*.
```

```
#Cross-sectional dependence can lead to bias in tests results (also called contemporaneous correlation).
```

```
pcdtest(fixed, test = c("lm"))
```

```
pcdtest(fixed, test = c("cd"))
```

```
#No evidence of cross-sectional dependence.
```

```
#####  
###
```

```
#4. Testing for serial correlation
```

```
#Serial correlation tests apply to macro panels with long time series.
```

```
#Not a problem in micro panels (with very few years).
```

```
#The null is that there is not serial correlation.
```

```
# Breusch-Godfrey/Wooldridge test
```

```
pbgtest(fixed)
```

#No evidence of serial correlation

###

#5. Testing for unit roots/stationarity

#The Dickey-Fuller test to check for stochastic trends.

#The null hypothesis is that the series has a unit root (i.e. non-stationary).

#If unit root is present you can take the first difference of the variable.

#Panel.set <- plm.data(pdata.frame(Panel), index = c("country", "year"))

library(tseries)

adf.test(Panel.set\$y, k=2)

#No evidence of a unit root.

#Testing for heteroskedasticity

#The null hypothesis for the Breusch-Pagan test is homoskedasticity.

#####

#Testing for heteroskedasticity

#The null hypothesis for the Breusch-Pagan test is homoskedasticity.

library(lmtest)

bptest(y ~ x1 + factor(country), data = Panel, studentize=F)

#No evidence of homoskedasticity

#If heteroskedasticity is detected you can use robust covariance matrix to account for it.

